图书情报工作
第 63 卷 第 23 期 2019 年 12 月

高校图书馆复杂网络构建与智慧化应用探索

■ 施国良[1] 谢泽宇[1] 杨小莉[2]
[1] 河海大学商学院 南京 211100 [2] 河海大学图书馆 南京 211100

摘要：[目的/意义]高校图书馆信息化水平高，但数据挖掘与智慧化水平有待提升。复杂网络以图数据库为存储和图查询的载体，对图结构数据进行统一组织和挖掘。图嵌入、图算法技术相较于传统机器学习方法能够充分挖掘图结构数据中的隐含联系。本研究运用复杂网络技术融合多源数据，探索图嵌入技术、图算法等图结构数据挖掘方法在提升图书馆智慧化水平中的作用。[方法/过程]首先基于可获取的数据进行数据特征分析与清洗；其次结合数据特征构建复杂网络概念模型，采用 Neo4j 批量导入技术实现网络构建和存储；最后探索图算法、图嵌入技术在图结构数据挖掘中的应用。[结果/结论]以图结构融合多源数据构建图书馆复杂网络，并以图数据库作为存储介质。图算法与图嵌入技术在在用户画像分析、精准推荐、智能问答等图书馆智能化应用等方面具有独特优势。

关键词：复杂网络 图数据库 图算法 图嵌入 智慧图书馆
分类号：G254
DOI：10.13266/j.issn.0252-3116.2019.23.012

2016 年《国家十三五规划纲要》以及 2017 年《政府工作报告》中都提出要大力发展人工智能[1]。近两年对于机器学习、深度学习、知识图谱等方面的研究愈发热烈，企业也迅速跟进。人工智能的迅速发展主要得益于数据收集和计算能力的提升以及深度学习、机器学习等数据挖掘方法的快速发展。在大数据和人工智能技术迅速发展的背景下，图书馆服务也由信息化向智能化转变[2]。综合运用人工智能技术，积极探索图书馆服务方面的智能化应用，有助于进一步满足读者需求、提高图书馆资源使用效率和服务效用。

目前，图书馆的大数据分析与智能化应用水平难以满足用户精准推荐、知识服务等需求[3]。本研究结合图书馆大数据，运用图数据库技术构建复杂网络，并探索图算法、图嵌入技术在用户画像、推荐系统与智能问答等图书馆智能化应用领域的作用。将复杂网络、图数据库、图算法、图嵌入等技术引入图书馆数据分析和挖掘中，有助于图书馆更好地服务于读者、提升资源使用效率。

1 研究现状

1.1 图书馆智能化服务领域

20 世纪 70 年代以来，智能化技术逐渐被应用于图书管理与服务中。1994 年，瑞典恩舍尔茨维克市立图书馆使用工业机器人对图书进行接收、分类和登记[4]。自助图书馆在中国各地落地，首都图书馆建设街区自助图书馆，应用 RFID 技术的小型自助图书馆于东莞市落地，可为市民提供全天图书借还服务[5]。智能参考咨询机器人如上海交通大学图书馆的“小交”，不仅可提供信息服务还能够与读者进行自主聊天[5]。虽然图书馆智能化应用水平还有待提升，但是这些应用对于图书资源使用效率和图书馆服务水平提升具有促进作用。

在大数据与智能化背景下，图书馆信息化程度不断提高，读者与系统的交互数据不断增长，新兴的数据挖掘方法不断引入图书馆服务与应用中，但是图书馆在信息化向智能化过渡中依然面临较大的挑战[6]。一方面，图书馆智慧服务需要专业化的人才以及大量的资金支持。专业人才擅长运用技术手段提升图书馆服务质量，图书馆信息化向智慧化转变的科研探索离不开资金支持。另一方面，图书馆在现有数据的收集与服务管理上也存在缺陷。数据是智能化应用的基石，图书馆在采集和管理大数据需要大量的软件、硬件设施作为保障，数据采集和使用需要遵守规范，需要注意

作者简介：施国良(ORCID:0000-0001-7585-640X)，副教授，博士；谢泽宇(ORCID:0000-0003-4330-0815)，硕士研究生，通讯作者：E-mail:18260062705@163.com；杨小莉(ORCID:0000-0001-8939-7232)，技术部主任，副研究馆员，硕士。
收稿日期：2019-03-03 修回日期：2019-07-02 本文起止页码：106-112 本文责任编辑：王传清

保护用户隐私、不得损害用户权利。图书馆作为知识、学习与交流中心，运用新兴的数据挖掘技术以及人工智能方法来提升服务水平和资源利用效率的价值巨大。本研究运用复杂网络和图数据库技术以图结构融合图书馆多元异构数据，结合图嵌入技术将实体向量化以进行智慧化应用，探索图书馆智能化道路中技术方面的可能途径。

1.2 复杂网络与图数据库领域

网络在自然界和人类社会中普遍存在，如社交网络、因特网、电力网络、航空网络等，复杂网络是对普遍存在的网络现象及其复杂性进行解释的一门学科，通常表现出小世界特性、层次特性、自组织特性等[7]。吴智勤等基于社交网络分析对高校图书馆进行用户画像分析研究[8]。冯蕾等提出在图书馆个性化推荐中融入复杂网络理论[9]。赵鹏等在自然语言处理领域运用复杂网络特征进行文档关键词抽取[10]。翟东升等在专利知识库构建中使用到图数据库[11]。李慧等在复杂网络环境中展开信任传递分析研究[12]。复杂网络理论及研究方法在图书馆数字资源整合、引文网、科研合作网等方面都有广泛的应用[13]。李德毅等发现在探索人的心智和思维机理的研究中复杂网络等理论的研究受到越来越多的关注。人们试图构建具有复杂网络特性的认知模型来体现人们的不确定认知过程[14]。图书馆是由各软硬件平台支撑的服务于读者的信息中心，也是用户与系统交互的关联数据中心。运用复杂网络思想和方法指导图书馆网络结构数据的挖掘与应用具有重要意义。

复杂网络依托于图数据库进行存储和应用，其数据存储结构和数据的查询方式都是以图论为基础的。Neo4j 作为应用最为广泛的企业级图数据库采用属性图模型实现，以节点、关系为基础数据结构能够描述绝大部分图的使用场景。基于图数据库技术的数据分析和挖掘在诸多领域已有成熟应用，如银行与保险业运用图数据库进行反欺诈模型研究；领英公司运用图数据库构建社交关系网络实现关系推荐。物流行业结合运输网络和图算法进行网络结构优化，降低运输成本等[15]。随着图数据库技术的不断进步，不仅实现关联网络的存储，还集成诸多经典图算法，如社区发现算法、中心性算法、路径规划等。图论算法将有助于对网络结构进行深入分析和研究。图嵌入技术将网络结构中的节点映射为低维稠密向量，进一步拓宽了图结构数据挖掘的手段[16]。综合运用复杂网络和图数据库融合多源数据，针对关联数据的分析和挖掘能够提升图书馆数据挖掘水平，从而提升智能化服务水平。

2 研究设计

复杂网络旨在将图书馆借阅数据整合为相互联系的数据网络，基于关联数据网络的数据挖掘相较于表存储结构的数据挖掘效果显著，可在网络结构分析的基础上提高智能化应用水平。图书馆复杂网络构建的流程为：①确定复杂网络涉及的数据边界和数据需求；②对各类抽取数据进行特征分析，结合复杂网络概念模型明确数据的清洗规则；；③按照确定的数据清洗规则对数据进行清洗，同时要注意读者隐私的保护；④应用 Neo4j 图数据库批量导入技术实现节点和关系的对应构建，⑤对批量导入完成的图数据库进行日志检查，并对数据库进行索引和查询优化；⑥在构建和优化的复杂网络上探索智能化应用。如图 1 所示：

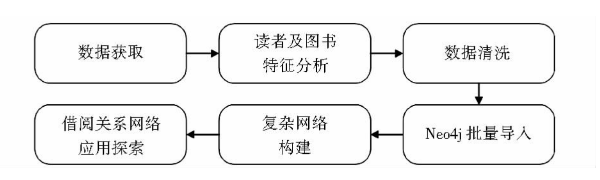


图 1　图书馆复杂网络构建流程

复杂网络由众多的节点和关系相互关联组成，底层的概念模型设计决定复杂网络的结构，网络结构的设计应综合考虑应用场景和图数据库性能。本研究结合图书信息数据、读者信息数据、读者借阅数据构建复杂网络。概念模型的设计主要是基于两个原则：其一是能够表征整个关系网络；其二是使图数据库子图查询的效率达到最优。Neo4j 图数据库的底层存储具有以下特征：①包含节点和关系；②节点和关系均有属性；③节点包含有一个或多个标签，关系仅有一种类型；④关系有向，从节点指向另一节点[17]。网络结构可抽象为读者、图书、作家等实体通过借阅关系、写作关系等构成的一张网络，复杂网络概念模型如图 2 所示：

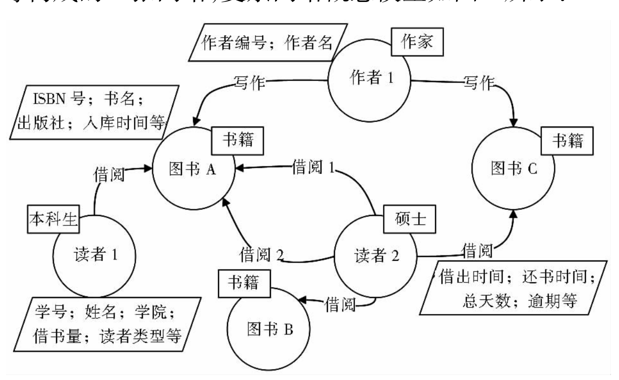


图 2　复杂网络概念模型

图书情报工作
第 63 卷 第 23 期 2019 年 12 月

3 复杂网络构建过程

3.1 数据来源

本研究的数据源是 A 图书馆图书借阅系统,数据的原始存储形式为 Oracle 关系型数据库。数据表是按照现有系统的使用流程进行设计,存在数据分离、冗余、类型多样等情况。本研究借助图数据库技术实现表存储结构数据向关联网络结构数据转化,并基于图算法以及图嵌入技术等针对图数据进行分析和挖掘。数据主要由 3 部分构成,读者信息主要包括读者姓名、专业、年级等;图书信息包括名称、分类号、作者、出版年等;图书借阅数据包括时间、浏览记录、借阅记录等。抽取年份范围为 2008 – 2018 年,共涉及 34 张数据表。

3.2 数据清洗

借阅网络概念模型的设计充分考虑到借阅网络的表征能力以及复杂网络在查询使用中的效率(见图 2)。Neo4j 图数据库本身的底层存储结构分为节点和关系两类,因此需要实现原始数据从表存储结构到图存储结构的转化。关系型数据库中表存储结构在描述某个复杂关联关系时会涉及多表操作以及深层查询,查询操作效率远低于图查询。复杂网络的节点与关系层次结构见图 3。整个复杂网络分为节点和关系,节点包含读者、图书、作者;关系包含借阅、写作。每个节点都有属性、主键以及标签,如读者节点中存储了学号、姓名、性别、学院、总借书量、信用情况、读者类型等属性,其中作为主键的学号是节点的唯一标识,节点可以拥有多个标签;每条关系同样可以存储属性以及类型。基于此实现表结构向图结构数据的转化,在关联关系分析中能够充分发挥图数据挖掘的优势。

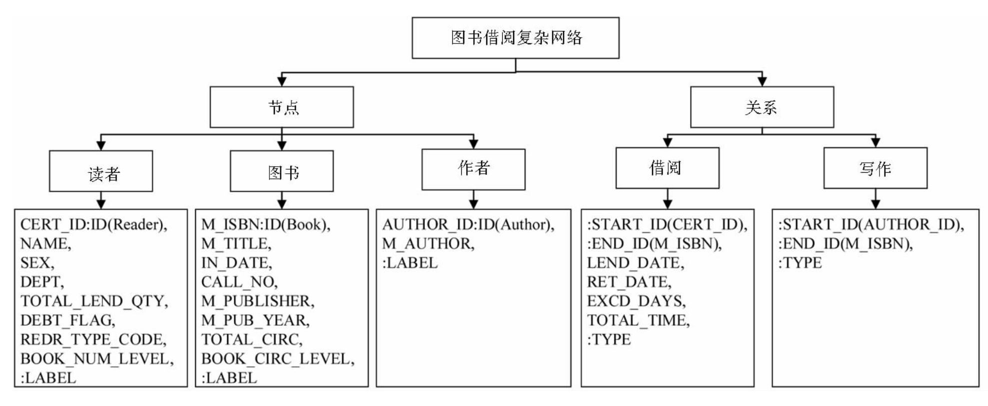


图 3 复杂网络节点、关系、属性

3.3 复杂网络的构建与存储

复杂网络是关系型数据库经过数据清洗之后,按照概念模型进行图结构转化而成的。图数据库数据导入有多种方式,包括 Cypher create 语句、Cypher load csv 语句、Java API、neo4j -import、neo4j -apocload 等,使用 neo4j-import 工具将整理好的 CSV 文件导入进库,速度能达到 12 万(节点 + 关系)/s,并且资源占用少。neo4j-import 工具批量导入命令见图 4。批量导入过程共耗时 19s 949ms,成功导入 715 682 节点、1 146 925 关系以及 2 861 027 属性。

```
neo4j-import
--multiline-fields=true
--bad-tolerance=1000
--into D:\software\neo4j-community-3.4.0\data\databases\Lib0818.db
--id-type string
--nodes E:\library\Final_lib\final\create\reader.csv
--nodes E:\library\Final_lib\final\create\book.csv
--nodes E:\library\Final_lib\final\create\auth.csv
--relationships E:\library\Final_lib\final\create\borrow.csv
--relationships E:\library\Final_lib\final\create\write.csv
```

图 4 neo4j-import 批量导入命令

批量导入完成后会在指定路径下生成 Lib0818.db 图数据库。Neo4j 图数据库集成可视化前端。开启 Neo4j 服务,修改配置文件中数据库的位置,然后在浏览器端输入网址(http://localhost:7474/browser/)能够连接指定数据库并实现查询和可视化操作。整个复杂网络由节点和边组成,边与节点相互关联构成联系紧密的关系网络。根据图 2 所述的概念模型进行复杂网络建模,对 Oracle 数据库中的表结构数据进行抽取、特征分析、清洗整理 CSV 文件,使用 neo4j-import 工具进行数据的批量导入,图数据库的查询和可视化操作在前端页面实现。见图 5。

3.4 图数据库优化

最大化地发挥图数据库的优势,需要对数据进行更加细致的划分,深度标签化;以图结构关联多源数据,基于关联数据网络能够挖掘更多的信息价值;数据量越大越能够挖掘出复杂网络中更隐晦的关联关系。图数据库作为复杂网络存储和应用的载体,需要保证

施国良，谢泽宇，杨小莉. 高校图书馆复杂网络构建与智慧化应用探索[J]. 图书情报工作,2019,63(23):106－112.

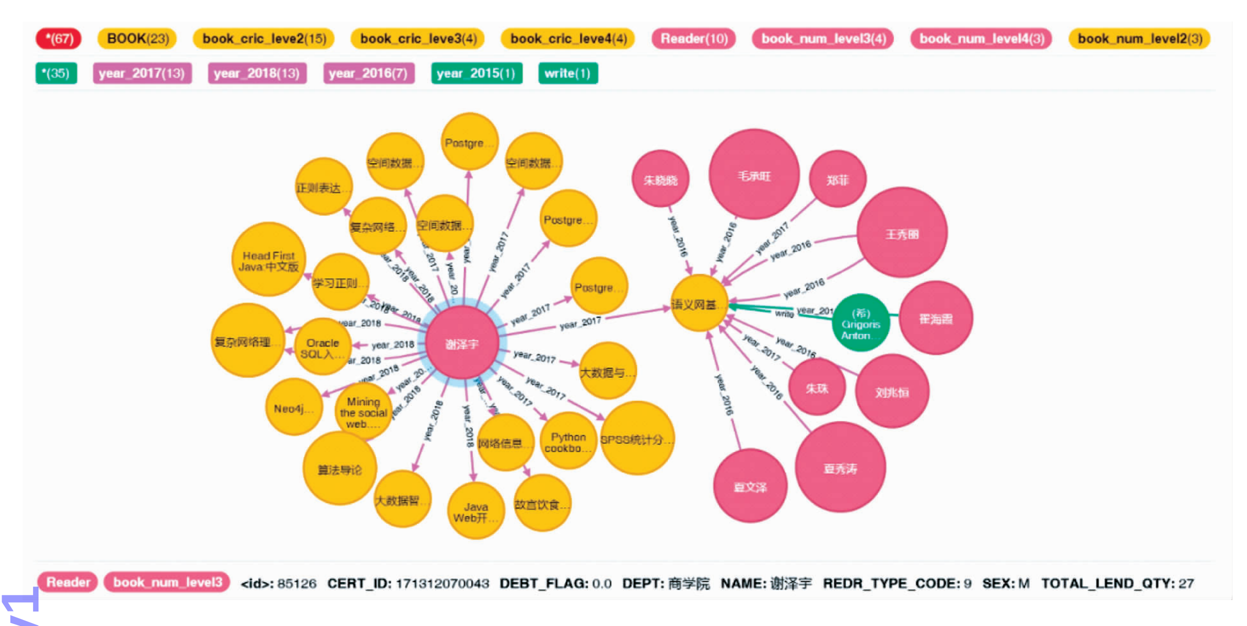


图 5　neo4j 图数据库前端页面

查询效率最优。Cypher 语言是 Neo4j 图数据库查询语言，不仅使得查询语句易于理解而且查询性能达到最优。查询优化方面可参考以下 5 点措施[15]：①深度标签化避免全局数据扫描；②引入索引和约束机制，复杂网络索引和约束 Cypher 语句见图 6；③避免笛卡尔积操作，避免查询非预期结果；④查询性能分析，使用 PROFILE 关键字监视查询的处理详细信息；⑤图模型优化，模型定义清晰、查询匹配迅速等。

```
create constraint on (a:BOOK) assert a.ISBN is unique
create constraint on (a:Reader) assert a.CERT_ID is unique
create constraint on (a:Author) assert a.AUTHOR_ID is unique
create index on :Author(NAME)
create index on :BOOK(TITLE)
create index on :Reader(NAME)
```

图 6　索引和约束 Cypher 语句

4　复杂网络应用探索

复杂网络实现了数据的融合，图算法与图嵌入将基于图数据进行分析和挖掘。复杂网络技术能够将多元异构数据进行整合，对来源不同、标准不同的数据进行统一管理，使数据挖掘起到“1＋1＞2”的效果。图结构数据的分析和挖掘主要可采用图算法与图嵌入技术两种思路。图算法如社区发现算法、PageRank 等；图嵌入技术将网络中的节点表示为低维稠密向量，使得机器学习、深度学习在网络结构挖掘中发挥作用。复杂网络与多源数据的关系类似于“超市货架与仓库”的关系，将多种数据用关联关系进行整合，充分发挥图数据引擎和图论算法以及图嵌入在网络结构数据挖掘中的作用。在图书馆大数据中运用复杂网络融合读者个人信息、浏览记录、读者借阅记录等多元数据进行综合分析能够增进系统对用户的了解，精准把握读者属性与偏好，能够充分运用到用户画像、个性化推荐以及知识问答等智慧化服务领域。

4.1　用户画像

用户画像是一种从海量数据中获取的、由用户信息构成的形象集合[19]。用户画像研究有利于图书馆精确把握读者需求，在提升图书馆服务质量、精准营销等活动中具有重要价值[20]。图书馆拥有大量用户、图书、交互等数据资源，能充分利用数据挖掘技术对数据进行组织、分析和挖掘，融合读者相关的多源数据构建关联数据之上的读者虚拟画像[21]。关联数据涵盖更全面的读者信息能够帮助精准提取用户标签，更加精准地刻画读者形象。

图书馆场景下用户画像存在以下问题：数据本身存在稀疏与分离，缺乏与外部数据进行互联互通；数据挖掘方法对单一数据源进行分析和挖掘，未融合多源数据充分利用数据的关联性。复杂网络将“数据孤岛”进行关联，融合多源数据进行统一组织与分析，能够挖掘单一数据源难以发现的多源数据间的关联关系。图 7 展示的是图书馆数据融合下读者用户画像的提取思路，多源数据不仅包括读者的个人属性信息、图书属性信息，还包括读者每次借阅行为产生的记录、浏览查询记录等。网络结构采用图算法进行分析和挖掘，在用户画像应用中读者个人与外部信息相互关联构成网络，运用社区发现算法可以实现用户的聚类，运用中心性算法（如 PageRank）可实现读者节点影响力

分析等。除了 Neo4j 图数据库内部集成相应的图算法，开源库（如 APOC）也能够辅助实现图算法。在数据关联的情况下运用图算法进行网络结构数据挖掘，能够更加准确把握读者之间的潜在关联与特征。

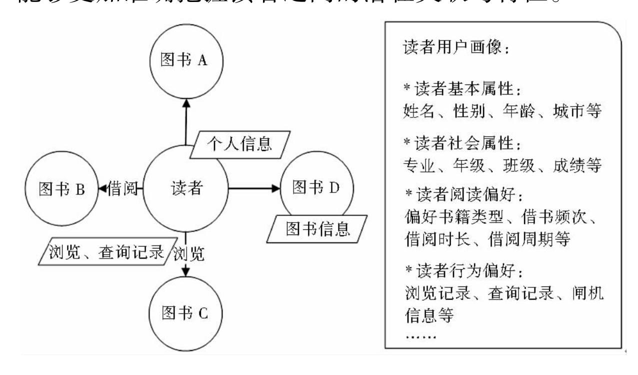


图 7　基于多源数据融合的读者用户画像分析

4.2　个性化推荐

传统的图书推荐系统以协同过滤以及基于内容的推荐为主，协同过滤算法是基于用户对物品的评分相似度进行推荐，喜爱相似物品的用户群体具有相近的偏好，然而大多高校图书馆缺乏读者对图书的打分数据仅包含借阅数据；由于读者的专业、学院以及图书的类别、分类号等属性相对稳定，基于内容的推荐难以实现动态的千人千面的推荐列表[22]。个性化推荐在不同的应用领域中有着不同的特点，高校图书馆图书推荐具有以下特点：推荐对象为图书，图书覆盖面广且以专业书籍居多；服务群体以师生为主，群体稳定且专业化程度高、易于聚类[23]。现阶段高校图书馆的推荐系统存在个性化程度低、数据收集、挖掘不充分的特点。

基于复杂网络的图书推荐服务策略有以下两种：①基于子图查询与自定义规则进行推荐；②基于图嵌入的方式进行实体向量化，根据实体相似性进行推荐。基于子图查询与规则的方式的优点在于实时动态推荐、缓解冷启动、规则灵活、推荐结果可解释性好以及用户体验好等。基于规则的复杂网络推荐策略见图 8，根据同一作者的书籍进行相似性推荐、基于图书的分类号分析实现基于内容的推荐、以及发现阅读偏好相近的其他读者的阅读列表实现聚类推荐等。但是人工定义规则难以适用于所有场景而且在大规模复杂网络中计算复杂度高，难以保证效率，而图嵌入技术将网络结构中的节点映射为低维稠密向量并基于向量化的实体进行推荐。目前很多科技公司的推荐系统都采用图嵌入技术进行大规模推荐[24]，各种图嵌入技术研究还在不断延伸，目前成熟的算法包含 DeepWalk[24]、node2Vec[25]、SDNE[26]、LINE[27] 等。使用图嵌入技术将节点向量化，通过节点的相似度计算就能够实现诸多应用，其中最典型的就是个性化推荐。实体嵌入极大地丰富了网络结构数据挖掘的方法，向量化表示的图结构能够运用更为灵活的机器学习以及深度学习方法而不仅仅局限于图算法。

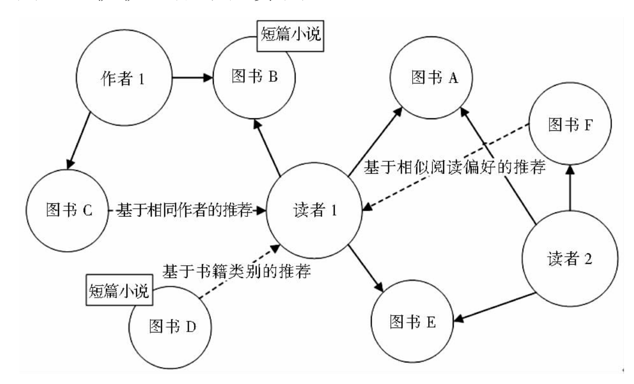


图 8　基于规则的复杂网络推荐策略

4.3　智能问答

在人工智能等技术日益发展的背景下，问答系统是图书馆智能化参考咨询问答服务的技术基础[28]。传统检索采用的是字符串匹配与排序算法相结合的方式将相关的信息网页展示给用户，智能问答是机器基于对问题的理解直接给出准确答案。知识图谱作为知识的载体成为问答系统的基础，除了通用型知识库如 DBpedia、YAGO、Freebase 等，领域知识库具有细粒度、高质量、推理复杂的特点。图书馆大数据构建领域知识库并以图数据库技术将知识存储起来，能够作为知识库辅助实现智能咨询。

目前问答系统实现方案主要有两种：①基于图匹配的方式。将自然语言中的语义关系表示成图，从而将自然语言问句转化为子图匹配的问题[29]。以复杂网络为知识库载体的基于图匹配实现的知识问答流程见图 9：首先对自然语言问句进行处理，抽取其中的实体；其次根据实体的类型匹配相应的问题模板，将原问题转化为子图查询语句；最后使用生成的查询语句对知识库进行查询，并将查询结果返回给用户。由于自然语言语法多样、表达复杂，基于子图匹配的方式实现准确的实体查找需要大量的标记数据和人工定义规则，不利于扩展。②基于向量化表示的问答对匹配方式。向量化表示节点不仅仅适用于推荐等应用，同样适用于知识问答。准备 QA 数据集用于训练节点和关系的向量化表示，将答案和问句映射到同一向量空间中。问句与回答的匹配仅需要计算向量化表示后的两个向量的相似度，而不需要考虑问句的语法与句法等。

此方法在通用知识图谱的智能问答相关的研究中已经取得较大进展[30－31]。知识库问答的另一种实现方式是图结构的向量化表示，如节点的向量化、路径的向量化以及子图的向量化，运用向量间的相似度计算实现知识问答。蕴含更多信息的向量化表示能够更加精准地回答问题。

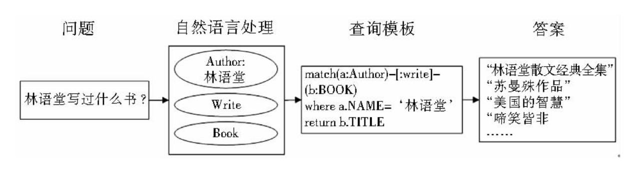


图9　基于图匹配的智能问答

5　结语

本研究基于A图书馆图书借阅以及读者和图书信息数据构建复杂网络，使用图数据库实现复杂网络的存储和查询，并基于此探索复杂网络数据挖掘在个性化推荐、用户画像、智能问答等智能化方面的应用。本文的主要贡献可以概括为：①复杂网络能够实现图书馆多平台大数据融合，将来源不同、标准不同的数据进行连接和整合，实现统一管理与组织，使数据挖掘起到“1＋1＞2”的效果。②运用图数据库技术实现复杂网络构造的完整流程。网络结构数据普遍存在，复杂网络构建技术具有普遍适用于关联数据网络，基于网络结构分析的应用都可以借鉴此构造流程。③探索图嵌入、图算法等网络结构数据挖掘方法在图书馆复杂网络分析中的应用。本研究将网络结构技术用于用户画像、推荐系统与智能问答等图书馆智能化应用领域中。积极探索复杂网络理论、图数据库技术、图嵌入、图算法等新型技术在图书馆大数据挖掘中的作用，对于推动图书馆服务智能化发展具有重要意义。

本研究的局限性在于：①复杂网络包括的数据仅仅只是图书馆大数据的一部分。若融合多平台图书馆大数据统一管理和挖掘，能够更好地服务于用户。②研究数据存在缺失、冗余等情况，同时考虑到用户隐私问题和用户权益不可侵犯，获取数据时对涉及用户隐私的字段例如：“身份证”“联系方式”等采用屏蔽以及脱敏处理，防止数据的滥用和过度挖掘。③本研究提出的图嵌入与图算法等网络结构挖掘方法缺乏实验验证，这也将是后续研究的方向。近两年，各学科、各科研院所以及企业在机器学习、深度学习、知识图谱等人工智能方面的研究正如火如荼地展开，图书馆也从信息化向智能化转变。当下图书馆信息化建设水平高，读者与图书馆交互数据充分，积极运用数据挖掘、人工智能技术，可以加速图书馆智能化转变的进程。

参考文献：

[1] 吴建中. 人工智能与图书馆[J]. 图书与情报, 2017(6):1－5.
[2] 李彩宁，毕新华，陈立军. 智慧图书馆服务模式及平台构建研究[J]. 图书馆, 2018(12):1－7.
[3] 布和宝力德. 人工智能技术在图书馆的应用、挑战及发展趋势[J]. 图书与情报, 2017(6):48－54.
[4] 刘晓民. 图书馆——自动化的新纪元[J]. 机器人技术与应用, 1997(1):7－8.
[5] 王展妮，张国亮. 图书馆机器人应用研究综述[J]. 大学图书馆学报, 2015,33(3):82－87.
[6] 杨九龙，阳玉堃，许碧涵. 人工智能在图书馆应用的理论逻辑、现实困境与路径展望[J]. 图书情报工作,2019,63(4):32－38.
[7] 汪小帆,李翔. 复杂网络理论及其应用[M]. 北京:清华大学出版社, 2006:18－46.
[8] 吴智勤，柳益君，李仁璞，等. 基于社交网络的高校图书馆用户画像构建研究[J]. 图书馆学研究, 2018(16):26－30.
[9] 冯蕾，张宇光，唐丽. 复杂网络理论在图书馆个性化推荐服务中的应用[J]. 情报理论与实践, 2009,32(2):69－71.
[10] 赵鹏，蔡庆生，王清毅，等. 一种基于复杂网络特征的中文文档关键词抽取算法[J]. 模式识别与人工智能, 2007,20(6):827－831.
[11] 翟东升，刘鹤，张杰，等. 基于图形数据库的专利语义知识库构建技术研究[J]. 现代图书情报技术, 2016(12):66－75.
[12] 李慧，马小平，施珺，等. 复杂网络环境下基于信任传递的推荐模型研究[J]. 自动化学报, 2018,44(2):363－376.
[13] 李晓瑛. 复杂网络理论及其在图书情报领域的应用研究[J]. 情报科学, 2016,34(10):95－98.
[14] 李德毅，刘常昱，杜鹢，等. 不确定性人工智能[J]. 软件学报, 2004(11):1583－1594.
[15] 张帜，庞国明，胡佳辉，等. Neo4j权威指南[M]. 北京:清华大学出版社, 2017:22－38.
[16] 刘知远，孙茂松，林衍凯，等. 知识表示学习研究进展[J]. 计算机研究与发展, 2016,53(2):247－261.
[17] KEMPER C. Managing your data in Neo4j[M]. Berkeley, CA: Apress, 2015:57－67.
[18] 刘海鸥，姚苏梅，黄文娜，等. 基于用户画像的图书馆大数据知识服务情境化推荐[J]. 图书馆学研究, 2018(24):57－63.
[19] 刘海鸥，孙晶晶，陈晶，等. 用户画像模型及其在图书馆领域中的应用[J]. 图书馆理论与实践, 2018(10):92－97.
[20] 陈添源. 高校移动图书馆用户画像构建实证[J]. 图书情报工作, 2018,62(7):38－46.
[21] 吕丹阳. 基于关联图谱的高校图书馆图书个性化推荐方法研究[D]. 武汉:华中科技大学, 2016.
[22] 李民，王颖纯，刘燕权. “211工程”高校图书馆馆藏资源推荐系统调查探析[J]. 图书情报工作, 2016,60(9):55－60.
[23] WANG J, HUANG P, ZHAO H, et al. Billion-scale commodity embedding for E-commerce recommendation in Alibaba[C] // Pro-

ceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. New York: ACM, 2018: 839 – 848.
[24] GROVER A, LESKOVEC J. node2vec: scalable feature learning for networks[C] // Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM, 2016: 855 – 864.
[25] WANG D, CUI P, ZHU W. Structural deep network embedding [C] // Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM, 2016:1225 – 1234.
[26] TANG J, QU M, WANG M, et al. LINE: large-scale information network embedding[C] // Proceedings of the 24th international conference on World Wide Web. New York: ACM, 2015:1067 – 1077
[27] 来云. 图书馆智能化咨询问答机器人系统设计与语料技术研究[J]. 现代情报, 2017,37(11):121 – 124.
[28] 沈奎林, 邵波, 赵华. 利用微信构建图书馆智能问答系统[J]. 图书馆学研究, 2015(8):75 – 80.
[29] BORDES A, WESTON J, USUNIER N. Open question answering with weakly supervised embedding models[C] // Joint European conference on machine learning and knowledge discovery in databases. Berlin, Heidelberg: Springer, 2014: 165 – 180.
[30] BORDES A, CHOPRA S, WESTON J. Question answering with subgraph embeddings[C] // Proceedings of the 2014 conference on empirical methods in natural language processing. Doha: EMNLP, 2014:615 – 620.

作者贡献说明:
施国良:提出研究命题、明确研究思路、修改论文;
谢泽宇:根据研究思路完成实验、撰写论文;
杨小莉:指导实验流程、完善研究思路、修改论文。

The Construction and Intelligent Applications of Complex Network in University Library

Shi Guoliang[1] Xie Zeyu[1] Yang Xiaoli[2]
[1] Business School, Hohai University, Nanjing 211100
[2] Hohai University Library, Nanjing 211100

Abstract: [**Purpose/significance**] The informatization level of university libraries is high, but the level of data mining and intelligence needs to be improved. The complex network uses graph database as the carrier of storage and graph query to organize and mine graph structure data. Compared with traditional machine learning methods, graph embedding and graph algorithm techniques can discover hidden connections in graph. This study uses complex network to integrate multi-source data and explores the role of graph data mining methods such as graph embedding and graph algorithms in improving library intelligence level. [**Method/process**] First of all, this study clarifies and analyzes the characteristics of the data based on the available data. Secondly, combined with the characteristics of data, construct a complex network conceptual model, and use Neo4j batch import technology to realize network construction and storage. Finally, explore the application of graph algorithm and graph embedding technology in graph structure data mining. [**Result/conclusion**] The multi-source data is combined with the graph structure to construct the complex network of the library, and the graph database is used as the storage medium. Graph algorithm and graph embedding technology have unique advantages in user image analysis, accurate recommendation, intelligent QA, and other intelligent applications of the library.

Keywords: complex network graph database graph algorithms graph embedding intelligent library